Traffic Flow Prediction with Relevance Vector Machine

W.-Z. Xiong
X.-M. Shen
H.-J. Li
Z. Shen (shenzhou.research@gmail.com)

Research Article

Keywords: Correlation vector machines, Traffic flow, Forecasting

Posted Date: June 1st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-547121/v1

Traffic Flow Prediction with Relevance Vector Machine

W.-Z. Xiong **X.-M. Shen** **H.-J. Li** **Z. Shen**

Abstract

Real-time prediction of traffic flow values in a short period of time is an important element in building a traffic management system. The uncertainty, complexity and nonlinearity of traffic flow data make it difficult to predict traffic flow in real time, and the accurate traffic flow prediction has been an urgent problem in the industry. Based on the research of scholars, a traffic flow prediction model based on the correlation vector machine method is constructed. The prediction accuracy of the correlation vector machine is better than that of the logistic regression and support vector machine methods, and the correlation vector machine method has the function of generating prediction error range for the actual traffic sequence data. The prediction results are very satisfactory, and the prediction speed is significantly faster than the other two models, which meets the requirement of real-time traffic flow prediction and is suitable for real-time online prediction, and the prediction accuracy of the used method is relatively high. The three-way comparison analysis shows that the traffic flow prediction by the correlation vector machine method can describe the nonlinear characteristics of traffic flow change more accurately, and the model performance and real-time performance are better. The case study shows that the traffic flow prediction model based on the correlation vector machine can improve the speed and accuracy of prediction, which is very suitable for traffic flow prediction estimation with real-time requirements, and provides a scientific method for real-time traffic flow measurement.

1 Introduction

With the continuous economic development, urbanization has accelerated, resulting in the continuous increase of traffic flow and congestion in the city, so intelligent transportation systems, which can effectively solve traffic problems, have emerged and are gaining attention [1–3]. The core of this system is how to accurately predict the traffic flow values that change from time to time, and the prediction results will directly affect the traffic planning and control problems [4–6]. The most obvious characteristic of traffic flow prediction is its short time, which is more susceptible to random factors than long-term prediction, and therefore has more obvious uncertainty and irregularity [2, 7].

In response to the above problems, many scholars began to study the traffic flow prediction problem. At present, the main methods used are: time series model, linear regression model, Kalman filter model, neural network model and learning algorithm support vector machine model [8–15]. Time series model mainly uses the linear characteristics of the data to predict the future data, the core lies in the use of linear characteristics, but it is difficult to accurately describe the traffic flow data with non-linear characteristics, the accuracy of prediction is biased; linear regression model mainly uses multiple linear regression model to analyze the traffic flow values, in the case of sufficient data, can be more traffic data values for faster prediction [16–18]. However, there are some disadvantages. The Kalman filter model is also a linear regression method, which is based on the further improvement of the multiple linear regression method, and can select the prediction factors according to the actual situation, and the prediction accuracy is improved compared with the multiple linear regression. The neural network algorithm is mainly based on the processing of risk, that is, minimizing

Preprint.

the empirical risk in prediction, but does not minimize the expected risk, so there are certain shortcomings, and the method requires some network structure, local minima, over-learning, etc., and the neural network needs to be based on a large number of The neural network requires a large number of training samples for data prediction. Support vector machine is a machine learning method and has been widely used. It solves the problems of overfitting, local minima and slow convergence in neural network algorithms, and is a new prediction method for traffic flow prediction. The relevance vector machine (RVM) is a new method based on the support vector machine, and has some characteristics, such as fewer relevance vectors and high sparsity compared with the support vector; using only the kernel parameters, the training time of the sample can be compressed; and the kernel function is not single, with a variety of selection space. The kernel function is not homogeneous and has multiple choice spaces. Therefore, in order to improve the accuracy of the traffic flow prediction method and reflect the real-time, this paper proposes a correlation vector machine based method for traffic flow prediction, and uses the method to conduct an example analysis to verify the prediction performance of the method.

2 Traffic Flow Prediction

2.1 Analysis of Traffic Flow Characteristics

Traffic flow is the sum of the number of vehicles that pass through a given road segment in a given unit of time. Traffic flow is characterized by uncertainty, real-time variability and high non-linearity. Considering the above characteristics, and the traffic flow on a certain road section, a certain time period and the traffic flow of the previous periods have a certain relationship, you can use the traffic flow data series of the previous periods to predict the traffic flow of a short period in the future, that is, based on historical traffic flow data and existing data, the application of some mathematical and statistical knowledge to build a prediction model. The traffic volume of a certain period of time in the future is predicted numerically. Let the traffic flow in the first t time periods be: $X_1, X_2, \ldots, X_t$, then the next time unit The traffic flow at the next time unit X_{t+1} can be described as: $X_{t+1} = f(X_1, X_2, \ldots, X_t)$.

2.2 Preprocessing of Sample Data

In the data training of the correlation vector machine model, if the data used in the training process has a large variability, it will slow down the learning speed of the model. In the training of the correlation vector machine model, if the data used in the training process has a large variability, it will slow down the learning speed of the model, which will affect the real-time prediction of the data. In order to speed up the training speed and convergence of the samples, it is necessary to make the initial traffic data collection In order to speed up the training speed and convergence of the samples, the initial traffic samples need to be further processed to improve the prediction speed and accuracy of the correlation vector machine model. In this paper In this paper, the traffic flow data is further processed by normalization method:

$$\hat{x}(i) = \frac{x(i) - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

where $\hat{x}(i)$ is the result of normalizing the data, $x_{\max}$ and $x_{\min}$ are the maximum and minimum values of the traffic data, respectively.

2.3 Construction of Prediction Model

The correlation vector machine is constructed based on the support vector machine, which is a new machine learning method based on Bayesian learning theory. The traffic flow observations of equal intervals in the same time period are represented as a time series S, $S = [s_1, s_2, ..., s_i, ..., s_M]$, assuming a highly nonlinear relationship between observation s_i and observations $s_{i-1}, s_{i-2}, ..., s_i - r$ (the r observations before s_i), the relationship can be expressed as follows:

$$s_i = \xi\left(s_{i-1}, s_{i-2}, \cdots, s_{i+}\right), i = r+1, r+2, \cdots, M \tag{2}$$

The above equation is based on the time series S of traffic flow data, and the sample set for traffic flow prediction in this paper is constructed as $\{X_i, \quad t_i\}_i^N = 1$ where $X_i = [s_i, s_{i+1}, \cdots, s_{i+r-1}]$,

T is the input sample for traffic prediction, $t_i = s_i + r$ is the output sample for traffic prediction, $N = M - r$ is the number of selected training samples.

Set the input target set as $\{X_n, t_n\}_i^N = 1$, where t is assumed to obey the independent distribution of Gaussian distribution function with expectation 0 and variance σ^2. As in the support vector machine, in the correlation vector machine, for a certain traffic flow training sample set $\{x_i, t_i\} \quad N_i = 1$ and assuming that the target The same prediction function as the support vector machine method is used here. Then the prediction equation of the correlation vector machine can be expressed as:

$$Y(X; W) = \sum_{i=1}^{N} W_i K(X, X_i) \tag{3}$$

where $K(X, X_i)$ is the radial basis function (RBF) and W_i is the regression coefficient. The construction of kernel function is mainly chosen The kernel function is a polynomial function obtained by combining a Gaussian function with a binomial function. When the output sets are independent of each other, the likelihood function of the whole sample set can be can be expressed as

$$p\left(t \mid W, \sigma^2\right) = \left(2\pi\sigma^2\right)^{N/2} \exp\left\{\frac{1}{2\sigma^2}\|t - \Phi W\|\right\} \tag{4}$$

where $W = [W_1, W_2, W_3, \cdots W_N]^T$ and $\Phi = \phi_i(X)_n$ is an $N \times N$ matrix with $\phi_i(X_n) = K(X_n, x_i)$. The prior distribution of W_i satisfies a Gaussian distribution with 0 as the mean and α_i^{-1} as the variance, denoted as

$$p(W \mid \alpha) = \prod_{i=0}^{N} N\left(W_i \mid 0, \alpha_i^{-1}\right) \tag{5}$$

where $\alpha = [\alpha_0, \alpha_1, \cdots \alpha_N]^T$ is denoted as the hyperparameter and α_i has the corresponding weight W_i. After defining the prior probability distribution and the likelihood distribution, the mathematical expression of the posterior distribution of the weight vector $W + i$ by Bayes' principle is expressed as follows:

$$p\left(t \mid W, \alpha, \sigma^2\right) = \left\{\begin{array}{l} = \frac{p\left(t \mid W, \sigma^2\right) p(W, \alpha)}{p\left(t \mid \alpha, \sigma^2\right)} \\ = (2\pi)^{-(N+1)/2} |\Sigma|^{+/2} \exp\left\{-\frac{1}{2}(W - m)^T \Sigma^{-1}(W - m)\right\} \end{array}\right\} \tag{6}$$

where $m = \sigma^{-2}\Sigma\Phi^T t$, $\Sigma = \left(\sigma^{-2}\Phi^{\mathrm{T}}\Phi + A\right)^{-1}$, and $A = \operatorname{diag}(\alpha_1, \alpha_2, \cdots, \alpha_N)$. The maximum likelihood function is obtained by using the integrated weights as follows:

$$p\left(t \mid \alpha, \sigma^2\right) = \left\{\begin{array}{l} = \int p\left(t \mid W, \sigma^2\right) p(W \mid \alpha) dW \\ = (2\pi)^{-v/2} |C|^{-1/2} \exp\left\{\frac{1}{2} t^T C^{-1} t\right\} \end{array}\right\} \tag{7}$$

where the covariance can be expressed as $C = \sigma^2 I + \Phi A^{-1} \Phi^T$. Taking the partial derivatives of α and σ^2, respectively, and making the partial derivatives equal to 0, we obtain:

$$\alpha_i^{\text{neis}} = \frac{\gamma_i}{m_i^2} \quad \left(\sigma^2\right)^{nev} = \frac{\|t - \Phi m\|}{N - \Sigma_i \gamma_i} \tag{8}$$

where m_i is the i-th element, $\gamma_i = 1 - \alpha_i \Sigma_{ii}$, and Σ_{ii} is the i-th diagonal element of matrix Σ.

The correlation vector machine continuously changes the values of m and Σ during the cyclic computation until the convergence requirement or the maximum number of iterations is reached. That is, in the internal cycle of the model In the internal loop of the model, most of the weights converge to zero, so that the terms of the kernel function matrix are not involved in the actual traffic prediction calculation to a large extent.

3 Case Studies

In order to verify the real-time performance and prediction accuracy of the RVM model, the traffic flow data of a detector on a road section in a city is used for comparison and validation. The detection period of the traffic data is 5 min. 5 samples of different time periods are used, and a total of 360 detections are made. The traffic flow data is predicted at one interval (5 min) in advance. The data are pre-processed before modeling, and the simulation program is prepared with MATLAB software to select reasonable parameters. Table 1 shows the evaluation results compared with other methods.

Because the correlation vector machine is a nonlinear prediction capability model, it can basically characterize the traffic flow variation law, and the RVM prediction results can better describe the nonlinear variation law of traffic flow. The prediction results can better describe the nonlinear variation law of traffic flow, and the prediction error can be reduced, and the prediction accuracy can be improved in real time. The prediction error is reduced, the prediction accuracy is improved in real time, and the accuracy is increased. In order to verify the accuracy and reliability of the RVM method, the prediction results of three methods, namely, logistic regression, support vector machine and correlation vector machine, are compared and analyzed. The prediction results of logistic regression, support vector machine and correlation vector machine are compared and analyzed as shown in Table 1.

Table 1: Comparison of the computational effort of different models

Model	Training Time (min)	Testing Time (min)	Accuracy
Logistic Regression	0.02	18.06	0.008
Support Vector Machine	0.02	2.46	0.050
Relevance Vector Machine	0.18	0.37	0.049

From the comparison parameters of the three types of methods in Table 1, it can be seen that the real-time traffic flow prediction of logistic regression and support vector machine is significantly higher than that of the correlation vector machine. The prediction accuracy of the correlation vector machine is better than the other two methods. The prediction accuracy of the correlation vector machine is better than the other two methods, and the correlation vector machine method has the function of generating the prediction error range. The correlation vector machine method has the ability to generate a prediction error range. For this set of actual traffic sequences, the proposed RVM-based traffic flow prediction model has the ability to generate a range of prediction errors. RVM-based traffic flow prediction model outperforms the logistic regression model and support vector machine in terms of overall prediction performance. The combined prediction performance of the RVM-based traffic prediction model is better than that of the logistic regression model and the support vector machine model. The prediction results are very satisfactory, and the prediction speed is significantly faster than the other two models. The prediction speed is significantly faster than the other two models, which can achieve the requirement of real-time traffic flow prediction. It is suitable for real-time online forecasting. Meanwhile, the prediction accuracy results in Table 1 show that the method used in this paper is very good. The comparative analysis of the three methods shows that the prediction of traffic flow using the correlation vector machine method is relatively accurate. The comparative analysis of the three methods shows that the traffic flow prediction using the correlation vector machine method can describe the nonlinear characteristics of traffic flow change more accurately, and the model performance and real-time performance are better, which is a more reliable method for traffic flow prediction. The model performance and real-time performance are better, and it is a more reliable method for traffic flow prediction. In summary, the correlation vector machine provides a traffic flow forecasting tool with high accuracy and good real-time performance. The correlation vector machine provides an accurate and real-time traffic forecasting tool.

4 Conclusions

In this paper, we propose a traffic flow prediction model based on RVM based on the special characteristics of traffic flow data. The RVM method solves the problems of long prediction time and poor generalization ability of neural network, and demonstrates its powerful small sample processing ability. The results of real-world data validation show that the traffic prediction by using the correlation vector machine The results of real-world data validation show that the prediction efficiency, accuracy and real-time performance are improved by using the correlation vector machine method to predict traffic flow. The comparison with other models shows that RVM is an effective traffic flow forecasting method with certain advantages in terms of prediction accuracy and time. RVM can output both predicted values and variance of predicted values, which is very suitable for traffic flow prediction estimation with real-time requirements. The RVM can output both forecast values and forecast variances, which is very suitable for traffic flow estimation with real-time requirements. In the next step, we can consider combining the correlation vector machine method with other methods. In the next step, we can consider combining the correlation vector machine method with other methods, i.e., building a combined prediction model based on the correlation vector machine, so that the prediction accuracy is higher and the prediction validity is more detailed. The research field of traffic flow prediction can be expanded in the comparison of continuous research.

References

[1] A. Perallos, U. Hernandez-Jayo, I. J. G. Zuazola, and E. Onieva, *Intelligent transport systems: technologies and applications*. John Wiley & Sons, 2015.

[2] J. J. Q. Yu, "Citywide traffic speed prediction: A geometric deep learning approach," *Knowledge-Based Systems*, p. 106592, Nov. 2020.

[3] J. J. Q. Yu and J. Gu, "Real-time traffic speed estimation with graph convolutional generative autoencoder," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 10, pp. 3940–3951, Oct. 2019.

[4] C. Zhang, J. J. Q. Yu, and Y. Liu, "Spatial-temporal graph attention networks: A deep learning approach for traffic forecasting," *IEEE Access*, vol. 7, pp. 166 246–166 256, 2019.

[5] X. Yin, G. Wu, J. Wei, Y. Shen, H. Qi, and B. Yin, "A comprehensive survey on traffic prediction," 2020, `arXiv:2004.08555 [eess.SP]`.

[6] A. Y. S. Lam, J. J. Q. Yu, Y. Hou, and V. O. K. Li, "Coordinated autonomous vehicle parking for vehicle-to-grid services: Formulation and distributed algorithm," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 4356–4366, Sep. 2018.

[7] P. Jaillet, J. Qi, and M. Sim, "Routing optimization under uncertainty," *Operations Research*, vol. 64, no. 1, pp. 186–200, 2016.

[8] Y. Liu, J. J. Q. Yu, J. Kang, D. Niyato, and S. Zhang, "Privacy-preserving traffic flow prediction: A federated learning approach," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7751–7763, Aug. 2020.

[9] Y. Shoukry, S. Mishra, Z. Luo, and S. Diggavi, "Sybil attack resilient traffic networks: a physics-based trust propagation approach," in *Proc. ACM/IEEE International Conference on Cyber-Physical Systems*, 2018, pp. 43–54.

[10] Y. Yu, J. J. Q. Yu, V. O. K. Li, and J. C. K. Lam, "A novel interpolation-SVT approach for recovering missing low-rank air quality data," *IEEE Access*, vol. 8, pp. 74 291–74 305, 2020.

[11] J. J. Q. Yu, "Semi-supervised deep ensemble learning for travel mode identification," *Transportation Research Part C: Emerging Technologies*, vol. 112, pp. 120–135, Mar. 2020.

[12] J. Guo, W. Huang, and B. M. Williams, "Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 50–64, 2014.

[13] J. J. Q. Yu, "Two-stage request scheduling for autonomous vehicle logistic system," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 5, pp. 1917–1929, May 2019.

[14] J. J. Q. Yu and A. Y. S. Lam, "Autonomous vehicle logistic system: Joint routing and charging strategy," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 7, pp. 2175–2187, Jul. 2018.

[15] H. R. Baghaee, D. Mlakić, S. Nikolovski, and T. Dragičević, "Support vector machine-based islanding and grid fault detection in active distribution networks," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, 2019.

[16] J. J. Q. Yu, "Sybil attack identification for crowdsourced navigation: A self-supervised deep learning approach," *IEEE Transactions on Intelligent Transportation Systems*, 2020, in press.

[17] C.-H. Wu, J.-M. Ho, and D. T. Lee, "Travel-time prediction with support vector regression," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 4, pp. 276–281, 2004.

[18] J. J. Q. Yu, D. J. Hill, V. O. K. Li, and Y. Hou, "Synchrophasor recovery and prediction: A graph-based deep learning approach," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 7348–7359, Oct. 2019.